cells

MDPI

Article

# Identification of miRNA signatures for kidney renal clear cell carcinoma using the tensor-decomposition method

**Ka-Lok Ng [1, 2] and Y-h Taguchi [3,*]**

[1] Department of Bioinformatics and Medical Engineering Asia University, Taichung, Taiwan; ppiddi@gmail.com
[2] Department of Medical Research, China Medical University Hospital China Medical University, Taiwan
[3] Department of Physics Chuo University, 1-13-27 Kasuga Bunky-ku, Tokyo 112-8551, Japan; tag@granular.com
* Correspondence: tag@granular.com

Received: date; Accepted: date; Published: date

**Abstract:** Cancer is a highly complex disease that is caused by multiple genetic factors at various levels. The use of microRNA (miRNA) and mRNA expression profiles is a useful method to identify prognostic biomarkers for cancer diagnosis. Traditional methods of identifying cancer prognostic markers may not be accurate. Tensor decomposition (TD) is a useful method uncovering the underlying low-dimensional structures in the tensor. We selected the kidney renal clear cell carcinoma (KIRC) in our analysis. We are able to identify the miRNA signatures and their significant correlated genes. These genes are involved in cancer-related pathways, and 23 genes are significantly correlated to the survival rates of the KIRC patients. This suggests that integrated analysis performed by TD-based unsupervised feature extraction technique is an effective strategy in identifying prognostic signatures in cancer studies.

**Keywords:** cancer biomarkers, diagnostic markers, prognostic markers, microRNA signatures, kidney cancer, Tensor decomposition

## 1. Introduction

Cancer is a highly complicated and heterogeneous disease. It is a result of loss of cell cycle control [1]. On the molecular biology level, the loss of control is a result of accumulation of genetic mutation, gene duplication [2] and aberrant epigenetic regulation [3, 4]. Genetic mutations involving activation of proto-oncogenes to oncogenes (OCG) and inactivation of tumor-suppressing genes (TSG). OCG and TSG may cause cancer by alternating the transcription factors (TF), such as the *p53* and *ras* oncoproteins, which in turn control expression of other genes. Gene duplication causes an elevated level of its protein product; thus favor the proliferation of the cancer cells. MicroRNAs (miRNAs) are a class of small non-coding RNAs that bind to the message RNA and induce either cleavage of mRNA or impede translation repression. Many studies have indicated that abnormal miRNA expression is associated with carcinogenesis [5]. MiRNAs induce cancers by acting as oncogenes (OCG) and tumor suppressor genes (TSG). A miRNA that targets the mRNA of a TSG would result in the loss of the protective effect of a TSG [5, 6]. Although there are many advancements in therapeutic strategies and diagnosis of the cancer disease, many patients are unable to get recover or the cancer recur after therapeutic treatment. The use of microRNA (miRNA) expression profiles is a useful method to identify prognostic biomarkers for cancer diagnosis. For instance, dysregulated miRNAs were identified in urothelial carcinoma of the bladder [7]. Recent works also suggested that miRNA could be used as a prognostic biomarker for patients with pancreatic adenocarcinoma [8, 9]. Also, by utilizing meta-analysis, it was reported that a panel of

eight-miRNA signature can serve as an effective marker for predicting the overall survival in bladder cancer patients [10]. In this study, we selected the kidney renal clear cell carcinoma (KIRC) in our analysis. KIRC is the most common cancer subtype of all renal malignant tumors (Zhang et al. 2013). It accounts for more than 70% of the cases. We noted that there are several studies have identified a few miRNA signatures that are associated with the overall survival of KIRC patients [11-13]. Those findings will be addressed in our discussion. Typical data structures in bioinformatics are difficult to analyze because there is a small number of samples with many variables. Supervised feature extraction are effective methods to reduce number of features. If supervised learning is applied, overfitting often can occur. Regularization (sparse modeling) try to minimize the number of features by restricting the sum of coefficients attributed to features and penalizes the usage of additional variables. The disadvantage of regularization is that we must select the values of parameters that balance the prediction accuracy and the number of variables. There are two major issues of the supervised feature extraction methods: (i) class labels may not always be true, and (ii) there may be more class labels present in the dataset. On the other hand, unsupervised methods such as principal component analysis (PCA) are often used to generate a smaller number of variables by the linear combination of original variables. The problem with this approach is that linear combination of many variables often prevent us from interpreting the newly generated variables. An unsupervised methodology that is suitable for the dimension reduction problems, which is tensor decomposition (TD)-based unsupervised feature extraction (FE) [14-27], can select smaller number of variables effectively and stably.

## 2. Materials and Methods

### *2.1. Tensor decomposition (TD) method*

The miRNAseq expression data for KIRC were retrieved from the TCGA Data Portal Research Network (https ://gdcportal.nci.nih.gov/).

TD is a natural extension of matrix factorization; it can be regarded as a generalization of the singular value decomposition (SVD) method. TD is a useful technique uncovering the underlying low-dimensional structures in the tensor. There are two popular tensor decomposition algorithms: the canonical polyadic decomposition (CPD) and the Tucker decomposition [28]. The rank decomposition method, CPD, is to express a tensor as the sum of a finite number of rank-one tensors. The Tucker decomposition decomposes a tensor into a so-called core tensor and multiple matrices.

TD-based unsupervised FE was applied to analyze the mRNA expression profiles and miRNA expression profiles. Let $x_{ij}^{(mRNA)}$ denotes the expression profiles of the $i$th mRNA ($i = 1, \ldots N$) of the $jth$ sample ( $j = 1, \ldots M$), whereas $x_{kj}^{(miRNA)}$ denotes the expression profiles of the $k$th miRNA ( $k = 1, \ldots K$) of the $j$th sample ( $j = 1, \ldots M$). Both $x_{ij}$ and $x_{kj}$ will be standardized such that they are associated with zero mean and unit variance. Next, we generated a case II type I tensor, that is,

$$x_{ijk} = x_{ij}^{(mRNA)} * x_{kj}^{(miRNA)} \tag{1}$$

$x_{ijk}$ is subjected to Tucker decomposition as follows:

$$x_{ijk} = \sum_{l_1=1}^{N} \sum_{l_2=1}^{M} \sum_{l_3=1}^{K} G(l_1, l_2, l_3)\, u_{l_1 i} u_{l_2 j} u_{l_3 k} \tag{2}$$

where $G \in R^{N\times M\times K}$ is the core tensor and $u_{l_1 i} \in R^{N\times N}$, $u_{l_2 j} \in R^{M\times M}$ and $u_{l_3 k} \in R^{K\times K}$ are singular value matrices that are orthogonal. Because Tucker decomposition is not unique, we have to specify how Tucker decomposition was derived. In particular, we chose higher-order singular value decomposition (HOSVD). Given that $x_{ijk}$ is too large to apply TD, we generated a case II type II tensor, which is given by:

$$x_{ik} = \sum_{j=1}^{M} x_{ijk} \tag{3}$$

By applying SVD, we can get $u_{l_1 i}$ and $u_{l_3 k}$ as

$$x_{ik} = \sum_{l=l_1=l_3=1}^{\min(N,K)} \lambda_l u_{l_1 i} u_{l_3 k} \quad (4)$$

Then we can also obtain two $u_{l_2 j}$ that correspond to miRNA and mRNA expression:

$$u_{l_1 j}^{mRNA} = \sum_{i=1}^{N} x_{ij} u_{l_1 i}, \; u_{l_3 j}^{miRNA} = \sum_{k=1}^{K} x_{kj} u_{l_3 k}, \quad (5)$$

Selection of genes can be determined by using the following quantities,

$$p_i = p_{\chi^2}\left[> \left(\frac{u_{l_1 i}}{\sigma_{l_1}}\right)^2\right], p_k = p_{\chi^2}\left[> \left(\frac{u_{l_3 k}}{\sigma_{l_3}}\right)^2\right] \quad (6)$$

where $p_{\chi^2}$[>x] is the cumulative probability that the argument is greater than $x$ in a $\chi^2$ distribution. $\sigma_{l_1}$ and $\sigma_{l_3}$ denote the standard deviations for $u_{l_1 i}$ and $u_{l_3 k}$, respectively. After the *p*-values are adjusted by means of the Benjamini–Hochberg (BH) criterion, miRNAs and mRNAs that are associated with adjusted *p*-values less than 0.01 are selected as those showing differences in expression between controls (normal tissues) and treated samples (tumors).

### *2.2 mRNA expression and miRNA expression*

Expression profiles of the mRNA and miRNA were retrieved from TCGA. The samples are composed of 253 kidney tumors and 71 normal kidney tissues (M=324). The number of mRNAs measured is N= 19536 while that of measured miRNAs is K=825.

### *2.3 Analysis of correlation between expression of the miRNAs and gene expression*

Correlations between $u_{l_1 j}^{mRNA}$ and $u_{l_3 j}^{miRNA}$ ($l_1 = l_3$ =2) were quantified by the Pearson's correlation coefficient (*PCC*). The *PCC* and p-vales were calculated by using the *corr.function* and *cor.test* function in R software, respectively.

### *2.4. Biological function analysis*

We evaluated the biological significance of the set of differentially expressed miRNAs and their correlated mRNAs. Biological annotations of the prognostic miRNAs and mRNAs were examined by employing the DIANA-miRPath [29] and MSigDB [30] databases respectively.

## 3. Results

We applied TD-based unsupervised FE to the KIRC dataset retrieved from TCGA. It was found that $u_{l_1 j}^{mRNA}$ and $u_{l_3 j}^{miRNA}$ ($l_1 = l_3$=2) are different between the normal and tumor samples. P-values computed by t test are $7.10 \times 10^{-39}$ for mRNA and $2.13 \times 10^{-71}$ for miRNA, respectively. In order to see if $u_{2j}^{mRNA}$ and $u_{2j}^{miRNA}$are significantly correlated, we computed Pearson's correlation coefficient between them. It is 0.9049376 (P = $1.63 \times 10^{-121}$). Thus they are highly correlated.

Figure 1. Scatter plot between $u_{l_1 j}^{mRNA}$ (vertical axis) and $u_{l_3 j}^{miRNA}$(horizontal axis). Black(red) open circles correspond to normal (tumor) tissues.

The results of the miRNA signatures and their significant correlated genes are shown in Table 1. A total of 11 miRNAs and 72 genes were identified. In order to see if these miRNAs and mRNAs are significantly correlated or not, we compute Pearson's correlation coefficient for all $11 \times 72 = 792$ pairs. Among them, 353 pairs are positively correlated and 358 pairs are negatively correlated (P-values are less than 0.01 after correcting with BH criterion). Thus, 89% of pairs are significantly correlated. Thus, we could successfully identified significantly correlated pairs of miRNAs and mRNAs.

**Table 1**. The results of the miRNA signatures and genes of the KIRC patients derived from the TD analysis.

| **miRNA ID** | | | | | |
|---|---|---|---|---|---|
| hsa-mir-210 | hsa-mir-891a | hsa-mir-155 | hsa-mir-200c | hsa-mir-141 | hsa-mir-508 |
| hsa-mir-122 | hsa-mir-514-3 | hsa-mir-514-1 | hsa-mir-514-2 | hsa-mir-184 | |
| **Gene symbol** | | | | | |
| ACTG1 | ADAM6 | AIF1L | ALDOA | ALDOB | ANGPTL4 |
| APLP2 | APP | AQP1 | AQP2 | ASS1 | ATP1A1 |
| ATP1B1 | ATP5A1 | ATP5B | B2M | C3 | C4A |
| C7 | CA12 | CCND1 | CD74 | CDH16 | COL4A1 |
| COL4A2 | CP | CYFIP2 | ENO1 | FN1 | FTL |
| GAPDH | GATM | GNB2L1 | GPX3 | HLA-A | HLA-B |
| HLA-C | HLA-DRA | HSD11B2 | HSP90AA1 | HSPA8 | IGFBP3 |
| IGFBP5 | ITM2B | KNG1 | LDHA | LDHB | LOC96610 |
| NDRG1 | NDUFA4L2 | NNMT | P4HB | PCK1 | PEBP1 |
| PLIN2 | PLVAP | PODXL | RGS5 | SERPINA1 | SLC12A1 |
| SLC12A3 | SOD2 | SPARC | SPP1 | TGFBI | TMBIM6 |
| TMSB10 | UBC | UMOD | VEGFA | VIM | VWF |

Next, in order to evaluate biological significance of selected mRNAs, the results of the top 10 oncogenic signatures of the 72 genes reported by MSigDB were summarized in Table 2.

**Table 2**. The top 10 oncogenic signatures of the 72 genes reported by the MSigDB. #Genes (K): the number of genes in each overexpressed gene set. # Genes in overlap (k): overlaps with genes selected by TD-based unsupervised FE method. .

| **Gene Set Name [# Genes (K)]** | **Description** | **#Genes in overlap (k)** | **p-value** | **FDR q-value** |
|---|---|---|---|---|
| CAMP_UP.V1_UP [200] | Genes up-regulated in primary thyrocyte cultures in response to cAMP signaling pathway activation by thyrotropin (TSH). | 7 | 9.97 e-8 | 1.88 e-5 |
| SNF5_DN.V1_DN [168] | Genes down-regulated in MEF cells (embryonic fibroblasts) with knockout of SNF5 [Gene ID=6598] gene. | 6 | 7.64 e-7 | 7.22 e-5 |
| ESC_V6.5_UP_ LATE.V1_UP [188] | Genes up-regulated during late stages of differentiation of embryoid bodies from V6.5 embryonic stem cells. | 6 | 1.47 e-6 | 9.27 e-5 |
| ESC_V6.5_UP_ EARLY.V1_DN [175] | Genes down-regulated during early stages of differentiation of embryoid bodies from V6.5 embryonic stem cells. | 5 | 1.98 e-5 | 8.54 e-4 |
| ESC_J1_UP_ LATE.V1_UP [189] | Genes up-regulated during late stages of differentiation of embryoid bodies from J1 embryonic stem cells. | 5 | 2.86 e-5 | 8.54 e-4 |
| SIRNA_EIF4GI_UP [95] | Genes up-regulated in MCF10A cells vs knockdown of EIF4G1 [Gene ID=1981] gene by RNAi. | 4 | 3.11 e-5 | 8.54 e-4 |
| P53_DN.V1_DN [193] | Genes down-regulated in NCI-60 panel of cell lines with mutated TP53 | 5 | 3.16 e-5 | 8.54 e-4 |

| | | | | |
|---|---|---|---|---|
| | [Gene ID=7157]. | | | |
| MEL18_DN.V1_UP [141] | Genes up-regulated in DAOY cells (medulloblastoma) upon knockdown of PCGF2 [Gene ID=7703] gene by RNAi. | 4 | 1.45 e-4 | 3.42 e-3 |
| LTE2_UP.V1_UP [188] | Genes up-regulated in MCF-7 cells (breast cancer) positive for ESR1 [Gene ID=2099] MCF-7 cells (breast cancer) and long-term adapted for estrogen-independent growth. | 4 | 4.33 e-4 | 8.51 e-3 |
| RPS14_DN.V1_UP [190] | Genes up-regulated in CD34+ hematopoietic progenitor cells after knockdown of RPS14 [Gene ID=6208] by RNAi. | 4 | 4.5 e-4 | 8.51 e-3 |

The results of the top 10 REACTOME pathways reported by MSigDB were summarized in Table 3.

**Table 3**. The top 10 oncogenic signatures of the 72 genes reported by the MSigDB. #Genes (K): the number of genes in each overexpressed gene set. # Genes in overlap (k): overlaps with genes selected by TD-based unsupervised FE method.

| **Gene Set Name [# Genes (K)]** | **Description** | **# Genes in overlap (k)** | **p-value** | **FDR q-value** |
|---|---|---|---|---|
| REACTOME_REGULATION_OF_INSULIN_LIKE_GR_GROWTH_FACTOR_IGF_TRANSPORT_AND_UPTAK TAKE_BY_INSULIN_LIKE_GROWTH_FACTOR_BIN BINDING_PROTEINS_IGFBPS [124] | Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs) | 12 | 9.03 e-18 | 1.35 e-14 |
| REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_NE_SYSTEM [856] | Cytokine Signaling in Immune system | 18 | 1.85 e-14 | 1.39 e-11 |
| REACTOME_RESPONSE_TO_ELEVATED_PLATELET LET_CYTOSOLIC_CA2PLUS [132] | Response to elevated platelet cytosolic Ca2+ | 9 | 3.42 e-12 | 1.71 e-9 |
| REACTOME_SIGNALING_BY_INTERLEUKINS [631] | Signaling by Interleukins | 13 | 1.53 e-10 | 4.86 e-8 |
| REACTOME_INNATE_IMMUNE_SYSTEM [1104] | Innate Immune System | 16 | 1.62 e-10 | 4.86 e-8 |
| REACTOME_PLATELET_ACTIVATION_SIGNALING ING_AND_AGGREGATION [260] | Platelet activation, signaling and aggregation | 9 | 1.45 e-9 | 3.63 e-7 |
| REACTOME_ENDOSOMAL_VACUOLAR_PATHWAY [11] | Endosomal/Vacuolar pathway | 4 | 3.63 e-9 | 7.78 e-7 |
| REACTOME_GLUCONEOGENESIS [34] | Gluconeogenesis | 5 | 5.22 e-9 | 9.79 e-7 |
| REACTOME_POST_TRANSLATIONAL_PROTEIN_MO | Post-translational protein modification | 16 | 6.56 e-9 | 1.09 e-6 |

| | | | | |
|---|---|---|---|---|
| _MODIFICATION [1429] | | | | |
| REACTOME_DISEASE [1075] | Disease | 14 | 1.02 e-8 | 1.53 e-6 |

These results suggested that the selceted 72 mRNAs are likely related to oncogenesis. In order to further confirm if these 72 mRNAs are related to kidny cancer, we check if these genes are related to survival rates (Table 4). Among 72 mRNAs, 23 mRNAs are significantly correlated to survival rates of kidney cancer patients. This also suggested the effectiveness of our analysis.

Table 4 Survival analysis of KIRC by using OncoLnc [31] (Kaplan plots are given in the supplementary matrials)

| gene | Cox Coeff. | p-Value | FDR Corrected | Rank | Median Expression | Mean Expression | Kaplan plot Low (%) | High (%) | p-value |
|---|---|---|---|---|---|---|---|---|---|
| VWF | -0.3 | 1.90E-04 | 1.41E-03 | 2253 | 23278.72 | 25958.34 | 50 | 50 | 3.99E-03 |
| VEGFA | 0.25 | 2.90E-03 | 1.19E-02 | 4064 | 31629.77 | 35072.27 | 70 | 30 | 2.32E-02 |
| TMBIM6 | -0.2 | 5.00E-03 | 1.82E-02 | 4583 | 27241.35 | 28733.19 | 40 | 60 | 1.34E-02 |
| PODXL | -0.4 | 8.00E-06 | 1.22E-04 | 1092 | 6659.17 | 7271.06 | 50 | 50 | 1.36E-06 |
| PLVAP | -0.2 | 7.10E-03 | 2.39E-02 | 4946 | 15470.76 | 17515.66 | 50 | 50 | 4.10E-04 |
| PLIN2 | -0.3 | 6.20E-04 | 3.56E-03 | 2902 | 18947.56 | 22839.08 | 50 | 50 | 1.71E-05 |
| PCK1 | -0.3 | 1.00E-04 | 8.58E-04 | 1931 | 1120.74 | 3037.73 | 50 | 50 | 7.84E-06 |
| NDRG1 | -0.2 | 1.20E-02 | 3.61E-02 | 5506 | 50127.14 | 51689.99 | 60 | 40 | 2.72E-02 |
| ITM2B | -0.3 | 6.00E-04 | 3.47E-03 | 2880 | 34751.8 | 36807.63 | 50 | 50 | 1.36E-02 |
| HSPA8 | -0.3 | 1.20E-03 | 5.90E-03 | 3363 | 17668.96 | 18139.95 | 40 | 60 | 1.04E-02 |
| HLA-DR A | -0.2 | 3.80E-03 | 1.46E-02 | 4304 | 29068.65 | 32924.27 | 20 | 80 | 4.22E-02 |
| GATM | -0.3 | 4.20E-04 | 2.61E-03 | 2683 | 5433.14 | 6800.94 | 50 | 50 | 3.09E-04 |
| CYFIP2 | -0.5 | 2.20E-09 | 4.32E-07 | 82 | 3482.26 | 4051.73 | 50 | 50 | 9.88E-08 |
| CDH16 | -0.2 | 4.40E-03 | 1.65E-02 | 4430 | 4093.23 | 4940.33 | 50 | 50 | 1.14E-03 |
| CCND1 | -0.2 | 3.00E-03 | 1.22E-02 | 4068 | 17278.68 | 19256.81 | 50 | 50 | 2.85E-04 |
| ATP5B | -0.2 | 1.10E-02 | 3.37E-02 | 5360 | 11450.7 | 13211.83 | 30 | 70 | 2.59E-03 |
| ATP5A1 | -0.2 | 2.20E-03 | 9.54E-03 | 3812 | 7988.24 | 9278.65 | 50 | 50 | 2.86E-02 |
| ATP1B1 | -0.3 | 1.50E-03 | 7.03E-03 | 3514 | 18741.07 | 21002.32 | 50 | 50 | 3.90E-02 |
| ATP1A1 | -0.3 | 4.90E-05 | 4.98E-04 | 1634 | 12917.72 | 15392.31 | 40 | 60 | 2.34E-02 |
| AQP1 | -0.3 | 4.30E-05 | 4.52E-04 | 1580 | 16717.87 | 19036.22 | 50 | 50 | 3.11E-08 |
| APP | -0.4 | 1.90E-05 | 2.36E-04 | 1329 | 32137.14 | 33051.3 | 50 | 50 | 1.33E-06 |
| ALDOB | -0.3 | 4.40E-05 | 4.61E-04 | 1587 | 467.22 | 3374.03 | 50 | 50 | 3.27E-06 |
| AIF1L | -0.2 | 1.50E-03 | 7.03E-03 | 3510 | 1984.01 | 2798.36 | 60 | 40 | 2.80E-02 |

We also evaluated 11 miRNAs by DIANA-mirpath (Table 5). It is obvous that many cancer related KEGG pathways are enrcihed. This also suggested the effectiveness of our strategy.

Table 5 Top 10 KEGG pathway enrichment of 11 miRNAs predicted by DIANA-mirpath (P-values are corrected). Full list can be obtained from http://snf-515788.vm.okeanos.grnet.gr/#mirnas=hsa-miR-210-3p;hsa-miR-210-5p;hsa-miR-891a-3p;hsa-miR-891a-5p;hsa-miR-200c-5p;hsa-miR-200c-5p;hsa-miR-141-5p;hsa-miR-141-3p;hsa-miR-122-3p;

hsa-miR-122-5p;hsa-miR-155-3p;hsa-miR-155-5p;hsa-miR-508-3p;hsa-miR-508-5p;hsa-miR-514a-3p; hsa-miR-514a-5p;hsa-miR-184&methods=Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase ;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase;Tarbase&selection=0

| KEGG pathway | p-value | #genes | #miRNAs |
|---|---|---|---|
| Chronic myeloid leukemia | 5.90E-08 | 39 | 6 |
| Proteoglycans in cancer | 3.67E-06 | 72 | 8 |
| Adherens junction | 2.58E-05 | 32 | 6 |
| TGF-beta signaling pathway | 2.58E-05 | 32 | 7 |
| Prostate cancer | 2.58E-05 | 43 | 7 |
| Pathways in cancer | 3.10E-05 | 128 | 10 |
| Pancreatic cancer | 3.94E-05 | 32 | 5 |
| Central carbon metabolism in cancer | 3.94E-05 | 29 | 5 |
| FoxO signaling pathway | 3.94E-05 | 54 | 7 |
| AMPK signaling pathway | 6.84E-05 | 51 | 5 |

## 4. Discussion

The top signature in Table 2 is a study of the cMAP signaling pathway. It has reported that targeting the cMAP pathway is an effective treatment of kidney cancer [32, 33]. The second signature in Table 2 is the study of the Snf5 gene expression profiles of the murine model (Mouse Embryonic Fibroblast (MEF) cells), which is closely resemble those of human SNF5-deficient rhabdoid tumors (pediatric soft tissue sarcoma that arises in the kidney, the liver, the peripheral nerves) [34]. Impairment of the SWI/SNF chromatin remodeling complex plays important function in the development and aggressiveness of clear cell renal cell carcinoma [35]. The sixth signature in Table 2 is a study of the effects of knockdown of the gene family, eukaryotic translation initiation factors (EIF), by RNAi in MCF10A cells. It has reported that EIF3b is a promising prognostic biomarker and a potential therapeutic target for patients with clear cell renal cell carcinoma [36], and EIF4GI is a target for cancer therapeutic [37].

The top pathway in Table 3 is the 'Pathway of regulation of IGF activity by IGFBP'. It was showed that the Insulin-like growth factors (IGFs) and insulin play a stimulatory role for renal cancer cells [38, 39]. Patients with IGF-1 receptor overexpression have 70 % increased risk of death [40], and has been shown to increase of kidney cancer risk in middle-aged male smokers [41]. The second pathway in Table 3 is the study of 'Cytokine Signaling in Immune system'. Cytokine is an important biomolecule that play essential roles in tumor formation [42] and it is a therapeutic target [43, 44]. The IL-6 cytokine family can serve as useful diagnostic and prognostic biomarkers. IL-6 is a potential target in cancer therapy [45, 46]. Ishibashi et al., reported that interleukin-6 (IL-6) suppresses the expression of the cytokine signaling-3 (SOCS3) gene, and IL-6 is associated with bad prognosis of kidney cancer patients [47].

It is rather obvious that Table 4 shows the significant relationship between 23 mRNAs and their kidney cancer survival rates. For some of the 23 genes, patients must be not divided into 50% and 50% based upon expression of considered genes in order to get significant p-vales for Kaplan plot. A majority of the mRNAs (15 out of 23 mRNAs) are associated with p-values less than 0.05 with 50% and 50% divisions based upon the amount of gene expression. Among the top ranked 10 KEGG pathways predicted by DIANA-mirpath (Table 5), six pathways; i.e. 1$^{st}$, 2$^{nd}$, 5$^{th}$, 6$^{th}$, 7$^{th}$ and 8$^{th}$ pathways, are directly related to cancers. For the rest four pathways, adherens junction is obviously related to cancers and the other three signaling pathways are also cancer related. Thus, we correctly identified 11 miRNAs that are cancer-related.

## 5. Conclusions

In this study, we applied TD-based unsupervised FE method to the KIRC miRNA expression and gene expression data. The TD-based method can identify miRNA signatures with differential expression between normal tissues and tumors as well as significant correlations between the gene expression data. Selected mRNAs and miRNAs are not only mutually correlated, but also are significantly related to various aspect of cancers. This suggests that integrated analysis performed by TD-based unsupervised FE is an effective strategy, in spite of its simplicity to identify biologically significant pairs of miRNAs and mRNAs, which is not easy by other strategies.

**Supplementary Materials** : Figure S1. The results of the Kaplan-Meier plots of the KIRC by using OncoLnc.

**Author Contributions**: Ka-Lok Ng foresee the research, prepared the data, writing—original draft preparation, review and editing. Y-h Taguchi performed the formal analysis, writing—original draft preparation, review and editing.

**Funding**: Dr. Ka-Lok Ng work was funded by the Ministry of Science and Technology, Taiwan (MOST), grant number MOST 108-2221-E-468-020, and also supported by the Asia University, grant numbers 107-asia-02 and 107-asia-09. It is also supported by Kakenhi 19H05270 and 17K00417.

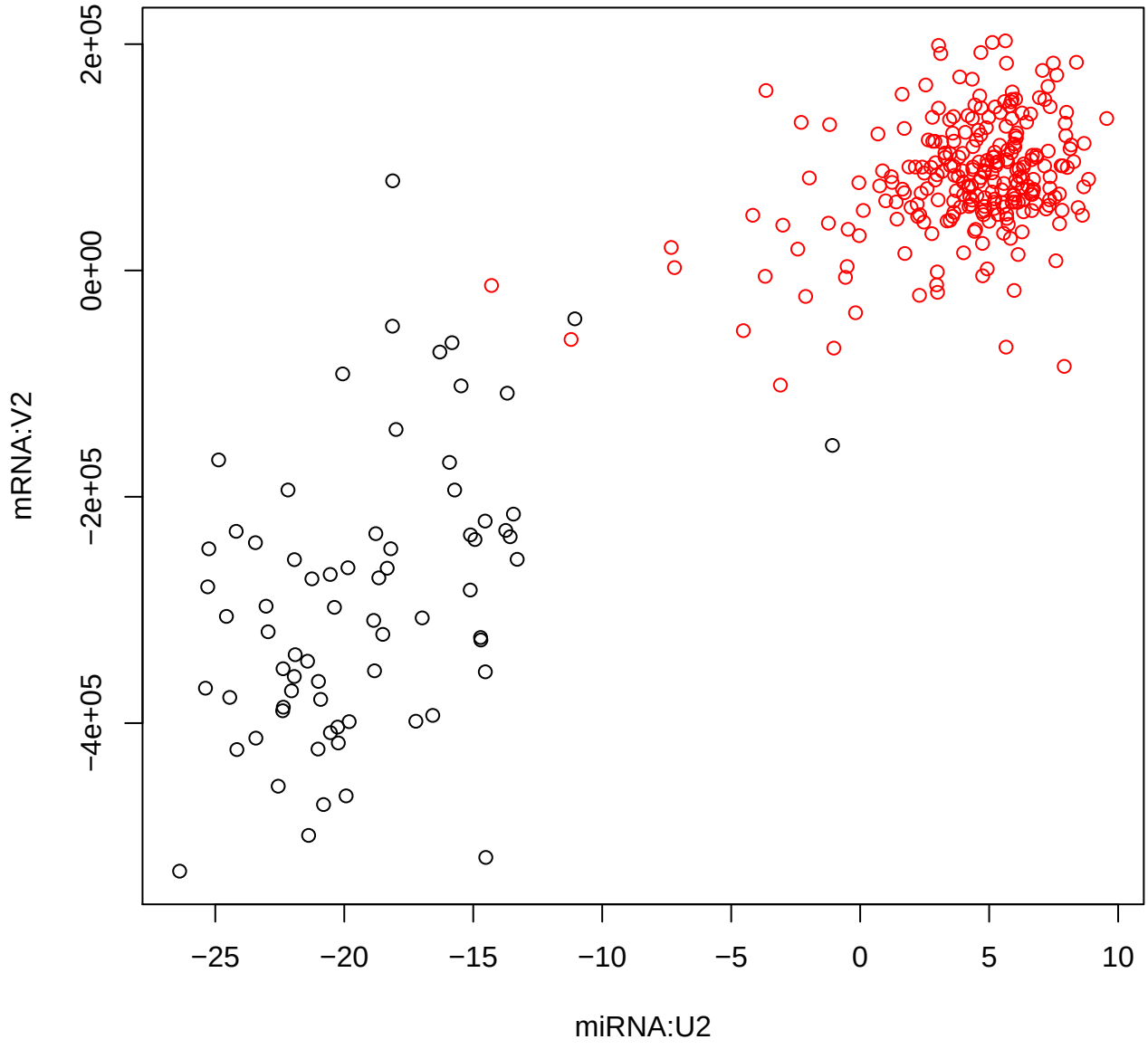
**Conflicts of Interest**: The authors declare no conflict of interest.

## References

1. Vargas-Rondon, N., V.E. Villegas, and M. Rondon-Lagos, *The Role of Chromosomal Instability in Cancer and Therapeutic Responses*. Cancers (Basel), 2017. **10**(1).
2. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. **144**(5): p. 646-74.
3. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts*. Nature, 1983. **301**(5895): p. 89-92.
4. Rouhi, A., et al., *MiRNAs, epigenetics, and cancer*. Mamm Genome, 2008. **19**(7-8): p. 517-25.
5. Medina, P.P. and F.J. Slack, *microRNAs and cancer: an overview*. Cell Cycle, 2008. **7**(16): p. 2485-92.
6. Zhang, W., J.E. Dahlberg, and W. Tam, *MicroRNAs in tumorigenesis: a primer*. Am J Pathol, 2007. **171**(3): p. 728-38.
7. Inamoto, T., et al., *A Panel of MicroRNA Signature as a Tool for Predicting Survival of Patients with Urothelial Carcinoma of the Bladder*. Dis Markers, 2018. **2018**: p. 5468672.
8. Shi, X.H., et al., *A Five-microRNA Signature for Survival Prognosis in Pancreatic Adenocarcinoma based on TCGA Data*. Sci Rep, 2018. **8**(1): p. 7638.
9. Yu, Y., X. Feng, and S. Cang, *A two-microRNA signature as a diagnostic and prognostic marker of pancreatic adenocarcinoma*. Cancer Manag Res, 2018. **10**: p. 1507-1515.
10. Zhou, H., et al., *A panel of eight-miRNA signature as a potential biomarker for predicting survival in bladder cancer*. J Exp Clin Cancer Res, 2015. **34**: p. 53.
11. Lokeshwar, S.D., et al., *Molecular Characterization of Renal Cell Carcinoma: A Potential Three-MicroRNA Prognostic Signature*. Cancer Epidemiol Biomarkers Prev, 2018. **27**(4): p. 464-472.
12. Luo, Y., et al., *Identification of a three-miRNA signature as a novel potential prognostic biomarker in patients with clear cell renal cell carcinoma*. J Cell Biochem, 2019. **120**(8): p. 13751-13764.
13. Xie, M., et al., *Identification and validation of a four-miRNA (miRNA-21-5p, miRNA-9-5p, miR-149-5p, and miRNA-30b-5p) prognosis signature in clear cell renal cell carcinoma*. Cancer Manag Res, 2018. **10**: p. 5759-5766.
14. Taguchi, Y−h. . *One-class Differential Expression Analysis using Tensor Decomposition-based Unsupervised Feature Extraction Applied to Integrated Analysis of Multiple Omics Data from 26 Lung Adenocarcinoma Cell Lines*. in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*. 2017.

15. Taguchi, Y−h. . and K. Ng. *[Regular Paper] Tensor Decomposition–Based Unsupervised Feature Extraction for Integrated Analysis of TCGA Data on MicroRNA Expression and Promoter Methylation of Genes in Ovarian Cancer*. in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*. 2018.
16. Taguchi, Y.-h., *Tensor Decomposition Based Unsupervised Feature Extraction Applied to Bioinformatics. Application of Omics, AI and Blockchain in Bioinformatics Research*. 2019: World Scientifc Publisher.
17. Taguchi, Y.-h., *Unsupervised Feature Extraction Applied to Bioinformatics: A PCA Based and TD Based Approach*. 2019: Springer International.
18. Taguchi, Y.-h., *Multiomics Data Analysis Using Tensor Decomposition Based Unsupervised Feature Extraction –Comparison with DIABLO–*. Intelligent Computing Theories and Application. ICIC 2019. Lecture Notes in Computer Science, ed. B.V. Huang DS., Premaratne P. Vol. 11643. 2019: Springer.
19. Taguchi, Y.-h. and T. Turki, *Neurological disorder drug discovery from gene expression with tensor decomposition*. Current Pharmaceutical Design, 2019: in press. .
20. Taguchi, Y.H., *Tensor decomposition-based unsupervised feature extraction identifies candidate genes that induce post-traumatic stress disorder-mediated heart diseases*. BMC Med Genomics, 2017. **10**(Suppl 4): p. 67.
21. Taguchi, Y.H., *Identification of candidate drugs using tensor-decomposition-based unsupervised feature extraction in integrated analysis of gene expression between diseases and DrugMatrix datasets*. Sci Rep, 2017. **7**(1): p. 13733.
22. Taguchi, Y.H., *Tensor decomposition-based unsupervised feature extraction applied to matrix products for multi-view data processing*. PLoS One, 2017. **12**(8): p. e0183933.
23. Taguchi, Y.H., *Tensor Decomposition-Based Unsupervised Feature Extraction Can Identify the Universal Nature of Sequence-Nonspecific Off-Target Regulation of mRNA Mediated by MicroRNA Transfection*. Cells, 2018. **7**(6).
24. Taguchi, Y.H., *Tensor decomposition-based and principal-component-analysis-based unsupervised feature extraction applied to the gene expression and methylation profiles in the brains of social insects with multiple castes*. BMC Bioinformatics, 2018. **19**(Suppl 4): p. 99.
25. Taguchi, Y.H., *Correction: Tensor decomposition-based unsupervised feature extraction applied to matrix products for multi-view data processing*. PLoS One, 2018. **13**(7): p. e0200451.
26. Taguchi, Y.H., *Drug candidate identification based on gene expression of treated cells using tensor decomposition-based unsupervised feature extraction for large-scale data*. BMC Bioinformatics, 2019. **19**(Suppl 13): p. 388.
27. Taguchi, Y.H. and T. Turki, *Tensor Decomposition-Based Unsupervised Feature Extraction Applied to Single-Cell Gene Expression Analysis*. Front Genet, 2019. **10**: p. 864.
28. Rabanser, S., O. Shchur, and S. Günnemann, *Introduction to Tensor Decompositions and their Applications in Machine Learning*. 2017.
29. Vlachos, I.S., et al., *DIANA-miRPath v3.0: deciphering microRNA function with experimental support*. Nucleic Acids Res, 2015. **43**(W1): p. W460-6.
30. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection*. Cell Syst, 2015. **1**(6): p. 417-425.
31. J., A., *OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs*. PeerJ Computer Science, 2016. **2**: p. e67.

32. Torres, V.E. and P.C. Harris, *Strategies targeting cAMP signaling in the treatment of polycystic kidney disease*. J Am Soc Nephrol, 2014. **25**(1): p. 18-32.
33. Piazzon, N., et al., *Bicc1 links the regulation of cAMP signaling in polycystic kidneys to microRNA-induced gene silencing*. J Mol Cell Biol, 2012. **4**(6): p. 398-408.
34. Isakoff, M.S., et al., *Inactivation of the Snf5 tumor suppressor stimulates cell cycle progression and cooperates with p53 loss in oncogenic transformation*. Proc Natl Acad Sci U S A, 2005. **102**(49): p. 17745-50.
35. Sarnowska, E., et al., *Evaluation of the role of downregulation of SNF5/INI1 core subunit of SWI/SNF complex in clear cell renal cell carcinoma development*. Am J Cancer Res, 2017. **7**(11): p. 2275-2289.
36. Zang, Y., et al., *Eukaryotic Translation Initiation Factor 3b is both a Promising Prognostic Biomarker and a Potential Therapeutic Target for Patients with Clear Cell Renal Cell Carcinoma*. J Cancer, 2017. **8**(15): p. 3049-3061.
37. Jaiswal, P.K., et al., *Eukaryotic Translation Initiation Factor 4 Gamma 1 (EIF4G1): a target for cancer therapeutic intervention?* Cancer Cell Int, 2019. **19**: p. 224.
38. Solarek, W., et al., *Insulin and insulin-like growth factors act as renal cell cancer intratumoral regulators*. J Cell Commun Signal, 2019. **13**(3): p. 381-394.
39. Braczkowski, R., et al., *Expression of insulin-like growth factor family genes in clear cell renal cell carcinoma*. Contemp Oncol (Pozn), 2016. **20**(2): p. 130-6.
40. Tracz, A.F., et al., *Insulin-like growth factor-1 signaling in renal cell carcinoma*. BMC Cancer, 2016. **16**: p. 453.
41. Major, J.M., et al., *Insulin-like growth factors and risk of kidney cancer in men*. Br J Cancer, 2010. **103**(1): p. 132-5.
42. Lee, M. and I. Rhee, *Cytokine Signaling in Tumor Progression*. Immune Netw, 2017. **17**(4): p. 214-227.
43. Doehn, C., et al., *Cytokine and vaccine therapy of kidney cancer*. Expert Rev Anticancer Ther, 2004. **4**(6): p. 1097-111.
44. Macleod, L.C., et al., *Trends in Metastatic Kidney Cancer Survival From the Cytokine to the Targeted Therapy Era*. Urology, 2015. **86**(2): p. 262-8.
45. Unver, N. and F. McAllister, *IL-6 family cytokines: Key inflammatory mediators as biomarkers and potential therapeutic targets*. Cytokine & Growth Factor Reviews, 2018. **41**: p. 10-17.
46. Kaminska, K., et al., *Interleukin-6 as an emerging regulator of renal cell cancer*. Urol Oncol, 2015. **33**(11): p. 476-85.
47. Ishibashi, K., et al., *Interleukin-6 induces drug resistance in renal cell carcinoma*. Fukushima J Med Sci, 2018. **64**(3): p. 103-110.


mRNA:V2
2e+05
0e+00
−2e+05
−4e+05
−25
−20
−15
−10
−5
0
5
10
miRNA:U2